

RECEIVED

SEP 04 2001

1631

TECH CENTER 1600/2900

RAW SEQUENCE LISTING

DATE: 08/08/2001

PATENT APPLICATION: US/09/322,067A

TIME: 11:04:09

Input Set : A:\10886-047001.txt

Output Set: N:\CRF3\08082001\I322067A.raw

4 <110> APPLICANT: Skolnick, Jeffrey
 5 Fetrow, Jacquelyn S.
 7 <120> TITLE OF INVENTION: METHODS AND SYSTEM FOR PREDICTING
 8 PROTEIN FUNCTION
 10 <130> FILE REFERENCE: 10886-047001
 12 <140> CURRENT APPLICATION NUMBER: 09/322,067A
 13 <141> CURRENT FILING DATE: 1999-05-27
 15 <150> PRIOR APPLICATION NUMBER: 60/099,300
 16 <151> PRIOR FILING DATE: 1998-08-25
 18 <150> PRIOR APPLICATION NUMBER: 60/120,311
 19 <151> PRIOR FILING DATE: 1999-02-16
 21 <160> NUMBER OF SEQ ID NOS: 12
 23 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 25 <210> SEQ ID NO: 1
 26 <211> LENGTH: 130
 27 <212> TYPE: PRT
 28 <213> ORGANISM: Haemophilus influenzae
 30 <400> SEQUENCE: 1
 31 Met Met Tyr Ser Lys Leu Leu Thr Leu Thr Thr Leu Leu Leu Pro Thr
 32 1 5 10 15
 33 Ala Leu Ala Leu Pro Ser Leu Val Glu Arg Ala Cys Asp Tyr Thr Cys
 34 20 25 30
 35 Gly Ser Asn Cys Tyr Ser Ser Ser Asp Val Ser Thr Ala Gln Ala Ala
 36 35 40 45
 37 Gly Tyr Gln Leu His Glu Asp Gly Glu Thr Val Gly Ser Asn Ser Tyr
 38 50 55 60
 39 Pro His Lys Tyr Asn Asn Tyr Glu Gly Phe Asp Phe Ser Val Ser Ser
 40 65 70 75 80
 41 Pro Tyr Tyr Glu Trp Pro Ile Leu Ser Ser Gly Asp Val Tyr Ser Gly
 42 85 90 95
 43 Gly Ser Pro Gly Ala Asp Arg Val Val Phe Asn Glu Asn Asn Gln Leu
 44 100 105 110
 45 Ala Gly Val Ile Thr His Thr Gly Ala Ser Gly Asn Asn Phe Val Glu
 46 115 120 125
 47 Cys Thr
 48 130
 50 <210> SEQ ID NO: 2
 51 <211> LENGTH: 107
 52 <212> TYPE: PRT
 53 <213> ORGANISM: Haemophilus influenzae
 55 <400> SEQUENCE: 2
 56 Gln Ser Ala Thr Thr Cys Gly Ser Thr Asn Tyr Ser Ala Ser Gln Val
 57 1 5 10 15
 58 Arg Ala Ala Ala Asn Ala Ala Cys Gln Tyr Tyr Gln Asn Asp Asp Ser
 59 20 25 30
 60 Ala Gly Ser Thr Thr Tyr Pro His Thr Tyr Asn Asn Tyr Glu Gly Phe
 61 35 40 45

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/322,067A

DATE: 08/08/2001

TIME: 11:04:09

Input Set : A:\10886-047001.txt

Output Set: N:\CRF3\08082001\I322067A.raw

```

62 Asp Phe Pro Val Asp Gly Pro Tyr Gln Glu Phe Pro Ile Lys Ser Gly
63      50                      55                      60
64 Gly Val Tyr Thr Gly Gly Ser Pro Gly Ala Asp Arg Val Val Ile Asn
65 65                      70                      75                      80
66 Thr Ile Asn Cys Glu Tyr Ala Gly Ala Ile Thr His Thr Gly Ala Ser
67                      85                      90                      95
68 Gly Asn Asn Phe Val Gly Cys Ser Gly Thr Asn
69                      100                      105
71 <210> SEQ ID NO: 3
72 <211> LENGTH: 105
73 <212> TYPE: PRT
74 <213> ORGANISM: Haemophilus influenzae
76 <400> SEQUENCE: 3
77 Glu Ser Cys Glu Tyr Thr Cys Gly Ser Thr Cys Tyr Trp Ser Ser Asp
78 1                      5                      10                      15
79 Val Ser Ala Ala Lys Ala Lys Gly Tyr Ser Leu Tyr Glu Ser Gly Asp
80                      20                      25                      30
81 Thr Ile Asp Asp Tyr Pro His Gly Tyr His Asp Tyr Glu Gly Phe Asp
82                      35                      40                      45
83 Phe Pro Val Ser Gly Thr Tyr Tyr Glu Tyr Pro Ile Met Ser Asp Tyr
84      50                      55                      60
85 Asp Val Tyr Thr Gly Gly Ser Pro Gly Ala Asp Arg Val Ile Phe Asn
86 65                      70                      75                      80
87 Gly Asp Asp Glu Leu Ala Gly Val Ile Thr His Thr Gly Ala Ser Gly
88                      85                      90                      95
89 Asp Asp Phe Val Ala Cys Ser Ser Ser
90                      100                      105
92 <210> SEQ ID NO: 4
93 <211> LENGTH: 113
94 <212> TYPE: PRT
95 <213> ORGANISM: Haemophilus influenzae
97 <400> SEQUENCE: 4
98 Cys Asn Ile Pro Glu Ser Thr Asn Cys Gly Gly Asn Val Tyr Ser Asn
99 1                      5                      10                      15
100 Asp Asp Ile Asn Thr Ala Ile Gln Gly Ala Leu Asp Asp Val Ala Arg
101                      20                      25                      30
102 Pro Asp Gly Asp Asn Tyr Pro His Gln Tyr Tyr Asp Glu Ala Ser Glu
103                      35                      40                      45
104 Asp Ile Thr Leu Cys Cys Gly Pro Gly Ser Trp Ser Glu Phe Pro Leu
105      50                      55                      60
106 Val Tyr Asn Gly Pro Tyr Tyr Ser Ser Arg Asp Asn Tyr Val Ser Pro
107 65                      70                      75                      80
108 Gly Pro Asp Arg Val Ile Tyr Gln Thr Asn Thr Gly Glu Phe Cys Ala
109                      85                      90                      95
110 Thr Val Thr His Thr Gly Ala Ala Ser Tyr Asp Gly Phe Thr Gln Cys
111                      100                      105                      110
112 Ser
115 <210> SEQ ID NO: 5
116 <211> LENGTH: 104

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/322,067A

DATE: 08/08/2001

TIME: 11:04:09

Input Set : A:\10886-047001.txt

Output Set: N:\CRF3\08082001\I322067A.raw

```

117 <212> TYPE: PRT
118 <213> ORGANISM: Haemophilus influenzae
120 <400> SEQUENCE: 5
121 Asp Cys Asp Tyr Thr Cys Gly Ser His Cys Tyr Ser Ala Ser Ala Val
122 1 5 10 15
123 Ser Asp Ala Gln Ser Ala Gly Tyr Gln Leu Glu Ser Ala Gly Gln Ser
124 20 25 30
125 Val Gly Arg Ser Arg Tyr Pro His Gln Tyr Arg Asn Tyr Glu Gly Phe
126 35 40 45
127 Asn Phe Pro Val Ser Gly Asn Tyr Tyr Glu Trp Pro Ile Leu Ser Ser
128 50 55 60
129 Gly Ser Thr Tyr Asn Gly Gly Gly Pro Gly Ala Asp Arg Val Val Phe
130 65 70 75 80
131 Asn Asp Asn Asp Glu Leu Ala Gly Leu Ile Thr His Thr Gly Ala Ser
132 85 90 95
133 Gly Asp Gly Phe Val Ala Cys Tyr
134 100
136 <210> SEQ ID NO: 6
137 <211> LENGTH: 102
138 <212> TYPE: PRT
139 <213> ORGANISM: Haemophilus influenzae
141 <400> SEQUENCE: 6
142 Ala Cys Ala Ala Thr Cys Gly Thr Val Cys Tyr Thr Ser Ser Ala Ile
143 1 5 10 15
144 Ser Ser Ala Gln Ala Ala Gly Tyr Asn Leu Tyr Ser Thr Asn Asp Asp
145 20 25 30
146 Val Ser Asn Tyr Pro His Glu Tyr His Asn Tyr Glu Gly Phe Asp Phe
147 35 40 45
148 Pro Val Ser Gly Thr Tyr Tyr Glu Phe Pro Ile Leu Lys Ser Gly Lys
149 50 55 60
150 Val Tyr Thr Gly Ser Ser Pro Gly Ala Asp Arg Val Ile Phe Asn Asp
151 65 70 75 80
152 Asp Asp Glu Leu Ala Gly Val Ile Thr His Thr Gly Ala Ser Gly Asn
153 85 90 95
154 Asn Phe Val Ala Cys Thr
155 100
157 <210> SEQ ID NO: 7
158 <211> LENGTH: 102
159 <212> TYPE: PRT
160 <213> ORGANISM: Haemophilus influenzae
162 <400> SEQUENCE: 7
163 Ala Cys Ala Ala Thr Cys Gly Ser Val Cys Tyr Thr Ser Ser Ala Ile
164 1 5 10 15
165 Ser Ala Ala Gln Glu Ala Gly Tyr Asp Leu Tyr Ser Ala Asn Asp Asp
166 20 25 30
167 Val Ser Asn Tyr Pro His Glu Tyr Arg Asn Tyr Glu Gly Phe Asp Phe
168 35 40 45
169 Pro Val Ser Gly Thr Tyr Tyr Glu Phe Pro Ile Leu Arg Ser Gly Ala
170 50 55 60

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/322,067A

DATE: 08/08/2001

TIME: 11:04:09

Input Set : A:\10886-047001.txt

Output Set: N:\CRF3\08082001\I322067A.raw

```

171 Val Tyr Ser Gly Asn Ser Pro Gly Ala Asp Arg Val Val Phe Asn Gly
172 65                               70                               75                               80
173 Asn Asp Gln Leu Ala Gly Val Ile Thr His Thr Gly Ala Ser Gly Asn
174                               85                               90                               95
175 Asn Phe Val Ala Cys Asp
176                               100
178 <210> SEQ ID NO: 8
179 <211> LENGTH: 104
180 <212> TYPE: PRT
181 <213> ORGANISM: Haemophilus influenzae
183 <400> SEQUENCE: 8
184 Ala Cys Met Tyr Ile Cys Gly Ser Val Cys Tyr Ser Ser Ser Ala Ile
185 1                               5                               10                               15
186 Ser Ala Ala Leu Asn Lys Gly Tyr Ser Tyr Tyr Glu Asp Gly Ala Thr
187                               20                               25                               30
188 Ala Gly Ser Ser Ser Tyr Pro His Arg Tyr Asn Asn Tyr Glu Gly Phe
189                               35                               40                               45
190 Asp Phe Pro Thr Ala Lys Pro Trp Tyr Glu Phe Pro Ile Leu Ser Ser
191 50                               55                               60
192 Gly Arg Val Tyr Thr Gly Gly Ser Pro Gly Ala Asp Arg Val Ile Phe
193 65                               70                               75                               80
194 Asp Ser His Gly Asn Leu Asp Met Leu Ile Thr His Asn Gly Ala Ser
195                               85                               90                               95
196 Gly Asn Asn Phe Val Ala Cys Asn
197                               100
199 <210> SEQ ID NO: 9
200 <211> LENGTH: 105
201 <212> TYPE: PRT
202 <213> ORGANISM: Haemophilus influenzae
204 <400> SEQUENCE: 9
205 Gln Gly Gly Val Ser Val Asn Cys Gly Gly Thr Tyr Tyr Ser Ser Thr
206 1                               5                               10                               15
207 Gln Val Asn Arg Ala Ile Asn Asn Ala Lys Ser Gly Gln Tyr Ser Ser
208                               20                               25                               30
209 Thr Gly Tyr Pro His Thr Tyr Asn Asn Tyr Glu Gly Phe Asp Phe Ser
210                               35                               40                               45
211 Asp Tyr Cys Asp Gly Pro Tyr Lys Glu Tyr Pro Leu Lys Thr Ser Ser
212 50                               55                               60
213 Ser Gly Tyr Thr Gly Gly Ser Pro Gly Ala Asp Arg Val Val Tyr Asp
214 65                               70                               75                               80
215 Ser Asn Asp Gly Thr Phe Cys Gly Ala Ile Thr His Thr Gly Ala Ser
216                               85                               90                               95
217 Gly Asn Asn Phe Val Gln Cys Ser Tyr
218                               100                               105
220 <210> SEQ ID NO: 10
221 <211> LENGTH: 177
222 <212> TYPE: PRT
223 <213> ORGANISM: Haemophilus influenzae
225 <400> SEQUENCE: 10

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/322,067A

DATE: 08/08/2001

TIME: 11:04:09

Input Set : A:\10886-047001.txt

Output Set: N:\CRF3\08082001\I322067A.raw

```

226 Met Val Ala Ile Lys Asn Leu Val Leu Val Ala Leu Thr Ala Val Thr
227 1 5 10 15
228 Ala Leu Ala Val Pro Ser Pro Leu Glu Ala Arg Ala Val Thr Trp Thr
229 20 25 30
230 Cys Leu Asn Asp Gln Lys Asn Pro Lys Thr Asn Lys Tyr Glu Thr Lys
231 35 40 45
232 Arg Leu Leu Tyr Asn Gln Asn Lys Ala Glu Ser Asn Ser His His Ala
233 50 55 60
234 Pro Leu Ser Asp Gly Lys Thr Gly Ser Ser Tyr Pro His Trp Phe Thr
235 65 70 75 80
236 Asn Gly Tyr Asp Gly Asp Gly Lys Leu Pro Lys Gly Arg Thr Pro Ile
237 85 90 95
238 Lys Phe Gly Lys Ser Asp Cys Asp Arg Pro Pro Lys His Ser Lys Asp
239 100 105 110
240 Gly Asn Gly Lys Thr Asp His Tyr Leu Leu Glu Phe Pro Thr Phe Pro
241 115 120 125
242 Asp Gly His Asp Tyr Lys Phe Asp Ser Lys Lys Pro Lys Glu Asn Pro
243 130 135 140
244 Gly Pro Ala Arg Val Ile Tyr Thr Tyr Pro Asn Lys Val Phe Cys Gly
245 145 150 155 160
246 Ile Ile Ala His Thr Lys Glu Asn Gln Gly Glu Leu Lys Leu Cys Ser
247 165 170 175
248 His
251 <210> SEQ ID NO: 11
252 <211> LENGTH: 177
253 <212> TYPE: PRT
254 <213> ORGANISM: Haemophilus influenzae
256 <400> SEQUENCE: 11
257 Met Val Ala Ile Lys Asn Leu Val Leu Val Ala Leu Thr Ala Val Thr
258 1 5 10 15
259 Ala Leu Ala Met Pro Ser Pro Leu Glu Glu Arg Ala Ala Thr Trp Ile
260 20 25 30
261 Cys Met Asn Glu Gln Lys Asn Pro Lys Thr Asn Lys Tyr Glu Asn Lys
262 35 40 45
263 Arg Leu Leu Tyr Asn Gln Asn Asn Ala Glu Ser Asn Ala His His Ala
264 50 55 60
265 Pro Leu Ser Asp Gly Lys Thr Gly Ser Ser Tyr Pro His Trp Phe Thr
266 65 70 75 80
267 Asn Gly Tyr Asp Gly Asp Gly Lys Ile Leu Lys Gly Arg Thr Pro Ile
268 85 90 95
269 Lys Trp Gly Asn Ser Asp Cys Asp Arg Pro Pro Lys His Ser Lys Asn
270 100 105 110
271 Gly Asp Gly Lys Asn Asp His Tyr Leu Leu Glu Phe Pro Thr Phe Pro
272 115 120 125
273 Asp Gly His Gln Tyr Asn Phe Asp Ser Lys Lys Pro Lys Glu Asp Pro
274 130 135 140
275 Gly Pro Ala Arg Val Ile Tyr Thr Tyr Pro Asn Lys Val Phe Cys Gly
276 145 150 155 160
277 Ile Val Ala His Thr Arg Glu Asn Gln Gly Asp Leu Lys Leu Cys Ser

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/322,067A

DATE: 08/08/2001

TIME: 11:04:10

Input Set : A:\10886-047001.txt

Output Set: N:\CRF3\08082001\I322067A.raw